

# SEQUENCE LISTING

<110> Kordyum, Vitaliy A.  
Chernykh, Svitlana I.  
Slavchenko, Iryna Yu.  
Vozianov, Oleksandr

<120> PHAGE-DEPENDENT SUPER PRODUCTION OF  
BIOLOGICALLY ACTIVE PROTEIN AND PEPTIDES

<130> PHAGE.006A

<150> 09/318,288

<151> 1999-05-25

<160> 11

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 630

<212> DNA

<213> Artificial Sequence

<220>

<223> This sequence was chemically synthesized based  
upon the amino acid sequence of human acidic  
fibroblast growth factor (155 amino acids) using  
codons which are used in highly expressed proteins  
from E. coli.

<221> CDS

<222> (122)...(590)

<400> 1

```
gcgtagagga tcgagatctc gatcccgcca aattaatacg actcactata ggggaattgt 60
gagcggataa caattcccct ctagaaataa ttttgtttaa ctttaagaag gagatataca 120
t atg gct gaa ggg gaa atc acc acc ttt aca gcg tta acg gag aaa ttt 169
Met Ala Glu Gly Glu Ile Thr Thr Phe Thr Ala Leu Thr Glu Lys Phe
      1             5             10             15
```

```
aac ctt ccg ccc ggg aat tac aaa aaa ccc aag ctt ctt tac tgc agt 217
Asn Leu Pro Gly Asn Tyr Lys Lys Pro Lys Leu Leu Tyr Cys Ser
      20             25             30
```

```
aac gga gga cac ttc ctg cga att ctg cca gat ggc aca gta gat ggg 265
Asn Gly Gly His Phe Leu Arg Ile Leu Pro Asp Gly Thr Val Asp Gly
      35             40             45
```

```
act cgc gat cgc tcc gac cag cac att cag ctg caa ctc tcg gcc gaa 313
Thr Arg Asp Arg Ser Asp Gln His Ile Gln Leu Gln Leu Ser Ala Glu
      50             55             60
```

```
agc gtt gga gag gtc tat atc aag tcg acg gag act ggc cag tac ctt 361
Ser Val Gly Glu Val Tyr Ile Lys Ser Thr Glu Thr Gly Gln Tyr Leu
```

65	70	75	80	
gcc atg gac acc gat ggg ctt ctg tat ggc tca cag acg cct aac gaa				409
Ala Met Asp Thr Asp Gly Leu Leu Tyr Gly Ser Gln Thr Pro Asn Glu				
	85	90	95	
gaa tgc ttg ttt cta gaa aga cta gaa gaa aac cat tac aac acg tac				457
Glu Cys Leu Phe Leu Glu Arg Leu Glu Glu Asn His Tyr Asn Thr Tyr				
	100	105	110	
ata tcg aaa aaa cat gca gag aag aac tgg ttt gta ggc ctt aaa aaa				505
Ile Ser Lys Lys His Ala Glu Lys Asn Trp Phe Val Gly Leu Lys Lys				
	115	120	125	
aat ggt tcc tgt aag cgt gga cca cgg act cac tat ggc caa aag gct				553
Asn Gly Ser Cys Lys Arg Gly Pro Arg Thr His Tyr Gly Gln Lys Ala				
	130	135	140	
atc ttg ttc ctg cca cta cca gtg agc tcc gac taa g gatccgaatt				600
Ile Leu Phe Leu Pro Leu Pro Val Ser Ser Asp *				
	145	150	155	
cgagctccgt cgacaagctt gcggccgcac				630
<210> 2				
<211> 155				
<212> PRT				
<213> Homo sapiens				
<400> 2				
Met Ala Glu Gly Glu Ile Thr Thr Phe Thr Ala Leu Thr Glu Lys Phe				
1 5 10 15				
Asn Leu Pro Pro Gly Asn Tyr Lys Lys Pro Lys Leu Leu Tyr Cys Ser				
20 25 30				
Asn Gly Gly His Phe Leu Arg Ile Leu Pro Asp Gly Thr Val Asp Gly				
35 40 45				
Thr Arg Asp Arg Ser Asp Gln His Ile Gln Leu Gln Leu Ser Ala Glu				
50 55 60				
Ser Val Gly Glu Val Tyr Ile Lys Ser Thr Glu Thr Gly Gln Tyr Leu				
65 70 75 80				
Ala Met Asp Thr Asp Gly Leu Leu Tyr Gly Ser Gln Thr Pro Asn Glu				
85 90 95				
Glu Cys Leu Phe Leu Glu Arg Leu Glu Glu Asn His Tyr Asn Thr Tyr				
100 105 110				
Ile Ser Lys Lys His Ala Glu Lys Asn Trp Phe Val Gly Leu Lys Lys				
115 120 125				
Asn Gly Ser Cys Lys Arg Gly Pro Arg Thr His Tyr Gly Gln Lys Ala				
130 135 140				
Ile Leu Phe Leu Pro Leu Pro Val Ser Ser Asp				
145 150 155				
<210> 3				
<211> 468				
<212> DNA				
<213> Homo sapiens				

<400> 3  
 atggctgaag gggaaatcac caccttcaca gccctgaccg agaagtttaa tctgcctcca 60  
 gggaattaca agaagcccaa actcctctac tgtagcaacg ggggccactt cctgaggatc 120  
 cttccggatg gcacagtgga tgggacaagg gacaggagcg accagcacat tcagctgcag 180  
 ctcagtgcgg aaagcgtggg ggaggtgtat ataaagagta ccgagactgg ccagtacttg 240  
 gccatggaca ccgacgggct tttatacggc tcacagacac caaatgagga atgtttgttc 300  
 ctggaaaggc tggaggagaa ccattacaac acctatatat ccaagaagca tgcagagaag 360  
 aattggtttg ttggcctcaa gaagaatggg agctgcaaac gcggtcctcg gactcactat 420  
 ggccagaaaag caatcttggt tctccccctg ccagtctctt ctgattaa 468

<210> 4  
 <211> 630  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> This sequence is a chemically synthesized sequence  
 encoding a 134 amino acid form of fibroblast  
 growth factor with alterations for preferred codon  
 usage in E. coli

<221> CDS  
 <222> (122)...(526)

<400> 4  
 gcgtagagga tcgagatctc gatcccgca aattaatacg actcactata ggggaattgt 60  
 gagcggataa caattcccct ctagaataaa ttttgtttaa ctttaagaag gagatataca 120  
 t atg aat tac aaa aaa ccc aag ctt ctt tac tgc agt aac gga gga cac 169  
 Met Asn Tyr Lys Lys Pro Lys Leu Leu Tyr Cys Ser Asn Gly Gly His  
 1 5 10 15  
 ttc ctg cga att ctg cca gat ggc aca gta gat ggg act cgc gat cgc 217  
 Phe Leu Arg Ile Leu Pro Asp Gly Thr Val Asp Gly Thr Arg Asp Arg  
 20 25 30  
 tcc gac cag cac att cag ctg caa ctc tcg gcc gaa agc gtt gga gag 265  
 Ser Asp Gln His Ile Gln Leu Gln Leu Ser Ala Glu Ser Val Gly Glu  
 35 40 45  
 gtc tat atc aag tcg acg gag act ggc cag tac ctt gcc atg gac acc 313  
 Val Tyr Ile Lys Ser Thr Glu Thr Gly Gln Tyr Leu Ala Met Asp Thr  
 50 55 60  
 gat ggg ctt ctg tat ggc tca cag acg cct aac gaa gaa tgc ttg ttt 361  
 Asp Gly Leu Leu Tyr Gly Ser Gln Thr Pro Asn Glu Glu Cys Leu Phe  
 65 70 75 80  
 cta gaa aga cta gaa gaa aac cat tac aac acg tac ata tcg aaa aaa 409  
 Leu Glu Arg Leu Glu Glu Asn His Tyr Asn Thr Tyr Ile Ser Lys Lys  
 85 90 95  
 cat gca gag aag aac tgg ttt gta ggc ctt aaa aaa aat ggt tcc tgt 457  
 His Ala Glu Lys Asn Trp Phe Val Gly Leu Lys Lys Asn Gly Ser Cys  
 100 105 110  
 aag cgt gga cca cgg act cac tat ggc caa aag gct atc ttg ttc ctg 505  
 Lys Arg Gly Pro Arg Thr His Tyr Gly Gln Lys Ala Ile Leu Phe Leu

115

120

125

cca cta cca gtg agc tcc gac taaggatccg aattcgagct ccgtcgacaa 556  
 Pro Leu Pro Val Ser Ser Asp  
 130 135

gcttgccggcc gcactcgagc accaccacca ccaccactga gatccggctg ctaacaaagc 616  
 ccgaaaggaa gctg 630

&lt;210&gt; 5

&lt;211&gt; 135

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

<223> Translated protein sequence for the chemically  
 synthesized 134 amino acid form of fibroblast  
 growth factor

&lt;400&gt; 5

Met Asn Tyr Lys Lys Pro Lys Leu Leu Tyr Cys Ser Asn Gly Gly His  
 1 5 10 15  
 Phe Leu Arg Ile Leu Pro Asp Gly Thr Val Asp Gly Thr Arg Asp Arg  
 20 25 30  
 Ser Asp Gln His Ile Gln Leu Gln Leu Ser Ala Glu Ser Val Gly Glu  
 35 40 45  
 Val Tyr Ile Lys Ser Thr Glu Thr Gly Gln Tyr Leu Ala Met Asp Thr  
 50 55 60  
 Asp Gly Leu Leu Tyr Gly Ser Gln Thr Pro Asn Glu Glu Cys Leu Phe  
 65 70 75 80  
 Leu Glu Arg Leu Glu Glu Asn His Tyr Asn Thr Tyr Ile Ser Lys Lys  
 85 90 95  
 His Ala Glu Lys Asn Trp Phe Val Gly Leu Lys Lys Asn Gly Ser Cys  
 100 105 110  
 Lys Arg Gly Pro Arg Thr His Tyr Gly Gln Lys Ala Ile Leu Phe Leu  
 115 120 125  
 Pro Leu Pro Val Ser Ser Asp  
 130 135

&lt;210&gt; 6

&lt;211&gt; 630

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> This sequence is a chemically synthesized sequence  
 encoding a 140 amino acid form of fibroblast  
 growth factor with alterations for preferred codon  
 usage in E. coli

&lt;221&gt; CDS

&lt;222&gt; (122)...(544)

&lt;400&gt; 6

gcgtagagga tcgagatctc gatcccgcca aattaatacg actcactata ggggaattgt 60  
 gagcggataa caattcccct ctagaataaa ttttgtttaa cttaagaag gagatataca 120

t atg ttt aac ctt ccg ccc ggg aat tac aaa aaa ccc aag ctt ctt tac 169  
 Met Phe Asn Leu Pro Pro Gly Asn Tyr Lys Lys Pro Lys Leu Leu Tyr  
 1 5 10 15

tgc agt aac gga gga cac ttc ctg cga att ctg cca gat ggc aca gta 217  
 Cys Ser Asn Gly Gly His Phe Leu Arg Ile Leu Pro Asp Gly Thr Val  
 20 25 30

gat ggg act cgc gat cgc tcc gac cag cac att cag ctg caa ctc tcg 265  
 Asp Gly Thr Arg Asp Arg Ser Asp Gln His Ile Gln Leu Gln Leu Ser  
 35 40 45

gcc gaa agc gtt gga gag gtc tat atc aag tcg acg gag act ggc cag 313  
 Ala Glu Ser Val Gly Glu Val Tyr Ile Lys Ser Thr Glu Thr Gly Gln  
 50 55 60

tac ctt gcc atg gac acc gat ggg ctt ctg tat ggc tca cag acg cct 361  
 Tyr Leu Ala Met Asp Thr Asp Gly Leu Leu Tyr Gly Ser Gln Thr Pro  
 65 70 75 80

aac gaa gaa tgc ttg ttt cta gaa aga cta gaa gaa aac cat tac aac 409  
 Asn Glu Glu Cys Leu Phe Leu Glu Arg Leu Glu Glu Asn His Tyr Asn  
 85 90 95

acg tac ata tcg aaa aaa cat gca gag aag aac tgg ttt gta ggc ctt 457  
 Thr Tyr Ile Ser Lys Lys His Ala Glu Lys Asn Trp Phe Val Gly Leu  
 100 105 110

aaa aaa aat ggt tcc tgt aag cgt gga cca cgg act cac tat ggc caa 505  
 Lys Lys Asn Gly Ser Cys Lys Arg Gly Pro Arg Thr His Tyr Gly Gln  
 115 120 125

aag gct atc ttg ttc ctg cca cta cca gtg agc tcc gac taaggatccg 554  
 Lys Ala Ile Leu Phe Leu Pro Leu Pro Val Ser Ser Asp  
 130 135 140

aattcgagct ccgtcgacaa gcttgcgggc gcaactcgagc accaccacca ccaccactga 614  
 gatccggctg ctaaca 630

<210> 7  
 <211> 141  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <223> Translated protein sequence for the chemically  
 synthesized 140 amino acid form of fibroblast  
 growth factor

<400> 7  
 Met Phe Asn Leu Pro Pro Gly Asn Tyr Lys Lys Pro Lys Leu Leu Tyr  
 1 5 10 15  
 Cys Ser Asn Gly Gly His Phe Leu Arg Ile Leu Pro Asp Gly Thr Val  
 20 25 30  
 Asp Gly Thr Arg Asp Arg Ser Asp Gln His Ile Gln Leu Gln Leu Ser  
 35 40 45  
 Ala Glu Ser Val Gly Glu Val Tyr Ile Lys Ser Thr Glu Thr Gly Gln

50                      55                      60  
 Tyr Leu Ala Met Asp Thr Asp Gly Leu Leu Tyr Gly Ser Gln Thr Pro  
 65                      70                      75                      80  
 Asn Glu Glu Cys Leu Phe Leu Glu Arg Leu Glu Glu Asn His Tyr Asn  
                     85                      90                      95  
 Thr Tyr Ile Ser Lys Lys His Ala Glu Lys Asn Trp Phe Val Gly Leu  
                     100                      105                      110  
 Lys Lys Asn Gly Ser Cys Lys Arg Gly Pro Arg Thr His Tyr Gly Gln  
                     115                      120                      125  
 Lys Ala Ile Leu Phe Leu Pro Leu Pro Val Ser Ser Asp  
                     130                      135                      140

<210> 8  
 <211> 1822  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <221> TATA\_signal  
 <222> (102)...(107)

<221> CDS  
 <222> (193)...(202)

<221> intron  
 <222> (203)...(458)

<221> CDS  
 <222> (459)...(619)

<221> intron  
 <222> (620)...(828)

<221> CDS  
 <222> (829)...(948)

<221> intron  
 <222> (949)...(1041)

<221> CDS  
 <222> (1042)...(1206)

<221> intron  
 <222> (1207)...(1459)

<221> CDS  
 <222> (1460)...(1654)

<223> Chemically synthesized sequence for Human Growth  
 Hormone using codons preferred for expression in  
 E. coli

<400> 8  
 ggagcttcta aattatccat tagcacaagc ccgtcagtgg ccccatgcat aaatgtacac 60  
 agaaacaggt gggggcaaca gtgggagaga agggggcagg gtataaaaag ggcccacaag 120  
 agaccggctc aaggatccca aggcccaact ccccgaaacca ctcagggtcc tgtggacgct 180

409420" 3766666

cacctagctg ca atg gct aca g gtaagcgccc ctaaaatccc tttgggcaca 232  
     Met Ala Thr  
     1  
  
 atgtgtcctg aggggagagg cagcgacctg tagatgggac gggggcacta accctcaggt 292  
 ttggggcttc tgaatgagta tgcctatgta agcccagtat ggccaatctc agaaagctcc 352  
 tggtccttgg agggatggag agagaaaaac aaacagctcc tggagcaggg agagtgtctg 412  
 cctcttgctc tccggctccc tctgttgccc tctggtttct ccccag gc tcc cgg acg 469  
                     Gly Ser Arg Thr  
                     5  
  
 tcc ctg ctc ctg gct ttt ggc ctg ctc tgc ctg ccc tgg ctt caa gag 517  
 Ser Leu Leu Leu Ala Phe Gly Leu Leu Cys Leu Pro Trp Leu Gln Glu  
     10                    15                    20  
  
 ggc agt gcc ttc cca acc att ccc tta tcc agg ctt ttt gac aac gct 565  
 Gly Ser Ala Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala  
     25                    30                    35  
  
 atg ctc cgc gcc cat cgt ctg cac cag ctg gcc ttt gac acc tac cag 613  
 Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln  
     40                    45                    50                    55  
  
 gag ttt gtaagctctt ggggaatggg tgcgcctcag ggggtggcagg aaggggtgac 669  
 Glu Phe  
  
 tttccccgc tgggaaataa gaggaggaga ctaaggagct cagggttttt cccgaagcga 729  
 aaatgcaggc agatgagcac acgctgagtg aggttcccag aaaagtaaca atgggagctg 789  
 gtctccagcg tagaccttgg tgggcgggtcc ttctctag gaa gaa gcc tat atc 843  
                                     Glu Glu Ala Tyr Ile  
                                     60  
  
 cca aag gaa cag aag tat tca ttc ctg cag aac ccc cag acc tcc ctc 891  
 Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu  
     65                    70                    75  
  
 tgt ttc tca gag tct att ccg aca ccc tcc aac agg gag gaa aca caa 939  
 Cys Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln  
     80                    85                    90  
  
 cag aaa tcc gtgagtggat gccttgaccc caggcgggga tgggggagac 988  
 Gln Lys Ser  
     95  
  
 ctgtagtcag agccccggg cagcacaggc caatgcccg tcttcccctg cag aac 1044  
                                     Asn  
  
 cta gag ctg ctc cgc atc tcc ctg ctg ctc atc cag tgc tgg ctg gag 1092  
 Leu Glu Leu Leu Arg Ile Ser Leu Leu Leu Ile Gln Ser Trp Leu Glu  
     100                    105                    110  
  
 ccc gtg cag ttc ctc agg agt gtc ttc gcc aac agc ctg gtg tac gcc 1140  
 Pro Val Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly  
     115                    120                    125                    130







80										85										90										
cag	acg	cct	aac	gaa	gaa	tgc	ttg	ttt	cta	gaa	aga	cta	gaa	gaa	aac	640														
Gln	Thr	Pro	Asn	Glu	Glu	Cys	Leu	Phe	Leu	Glu	Arg	Leu	Glu	Glu	Asn															
			95					100						105																
cat	tac	aac	acg	tac	ata	tcg	aaa	aaa	cat	gca	gag	aag	aac	tggt	ttt	688														
His	Tyr	Asn	Thr	Tyr	Ile	Ser	Lys	Lys	His	Ala	Glu	Lys	Asn	Trp	Phe															
		110					115					120																		
gta	ggc	ctt	aaa	aaa	aat	ggt	tcc	tgt	aag	cgt	gga	cca	cgg	act	cac	736														
Val	Gly	Leu	Lys	Lys	Asn	Gly	Ser	Cys	Lys	Arg	Gly	Pro	Arg	Thr	His															
		125				130					135																			
tat	ggc	caa	aag	gct	atc	ttg	ttc	ctg	cca	cta	cca	gtg	agc	tcc	gac	784														
Tyr	Gly	Gln	Lys	Ala	Ile	Leu	Phe	Leu	Pro	Leu	Pro	Val	Ser	Ser	Asp															
140					145				150						155															
taaggatccg aattcgagct ccgtcgacaa gcttgccggcc gcaactcgagc accaccacca 844																														
ccaccactga gatccggctg ctaacaaagc ccgaaaggaa gctgagttgg ctgctgccac 904																														
cgctgagcaa taactagcat aaccoccttg ggccctctaaa cgggtcttga ggggtttttt 964																														
gctgaaagga ggaactatat ccggat 990																														
<210> 11																														
<211> 155																														
<212> PRT																														
<213> Homo sapiens																														
<220>																														
<223> Translated protein sequence for the chemically synthesized human interferon alpha-2b																														
<400> 11																														
Met	Ala	Glu	Gly	Glu	Ile	Thr	Thr	Phe	Thr	Ala	Leu	Thr	Glu	Lys	Phe															
1				5				10						15																
Asn	Leu	Pro	Pro	Gly	Asn	Tyr	Lys	Lys	Pro	Lys	Leu	Leu	Tyr	Cys	Ser															
			20				25						30																	
Asn	Gly	Gly	His	Phe	Leu	Arg	Ile	Leu	Pro	Asp	Gly	Thr	Val	Asp	Gly															
		35				40					45																			
Thr	Arg	Asp	Arg	Ser	Asp	Gln	His	Ile	Gln	Leu	Gln	Leu	Ser	Ala	Glu															
	50				55					60																				
Ser	Val	Gly	Glu	Val	Tyr	Ile	Lys	Ser	Thr	Glu	Thr	Gly	Gln	Tyr	Leu															
65				70					75					80																
Ala	Met	Asp	Thr	Asp	Gly	Leu	Leu	Tyr	Gly	Ser	Gln	Thr	Pro	Asn	Glu															
			85					90						95																
Glu	Cys	Leu	Phe	Leu	Glu	Arg	Leu	Glu	Glu	Asn	His	Tyr	Asn	Thr	Tyr															
			100				105						110																	
Ile	Ser	Lys	Lys	His	Ala	Glu	Lys	Asn	Trp	Phe	Val	Gly	Leu	Lys	Lys															
		115				120						125																		
Asn	Gly	Ser	Cys	Lys	Arg	Gly	Pro	Arg	Thr	His	Tyr	Gly	Gln	Lys	Ala															
	130				135						140																			
Ile	Leu	Phe	Leu	Pro	Leu	Pro	Val	Ser	Ser	Asp																				
145					150					155																				